

日 本 国 特 許 庁

JAPAN PATENT OFFICE

27.12.02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年12月27日

REC'D 03 MAR 2003

WIPO

PCT

出 願 番 号

Application Number:

特願2001-398570

[ST.10/C]:

[JP2001-398570]

出 願 人

Applicant(s):

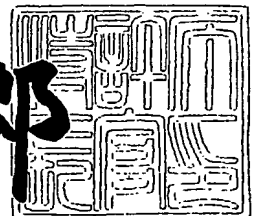
セレスター・レキシコ・サイエンス株式会社

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

2003年 2月12日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3005956

【書類名】 特許願

【整理番号】 PCLA-13312

【提出日】 平成13年12月27日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/00
G06F 17/30

【発明者】

【住所又は居所】 千葉県千葉市美浜区中瀬1丁目3番地 幕張テクノガーデンD17 セレスター・レキシコ・サイエンシズ株式会社内

【氏名】 齋藤 静司

【特許出願人】

【識別番号】 500520628

【氏名又は名称】 セレスター・レキシコ・サイエンシズ株式会社

【代理人】

【識別番号】 100089118

【弁理士】

【氏名又は名称】 酒井 宏明

【選任した代理人】

【識別番号】 100113103

【弁理士】

【氏名又は名称】 香島 拓也

【手数料の表示】

【予納台帳番号】 036711

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0117745

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体

【特許請求の範囲】

【請求項 1】 目的のタンパク質の一次配列情報を入力する入力手段と、
タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行手段と、

上記断片構造予測プログラム実行手段による上記断片構造予測プログラムの断片構造予測結果を比較する予測結果比較手段と、

上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、

上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段と、

を備えたことを特徴とする相互作用部位予測装置。

【請求項 2】 目的のタンパク質の一次配列情報を入力する入力手段と、
上記目的のタンパク質の立体構造データを取得する立体構造データ取得手段と

タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行手段と、

上記断片構造予測プログラム実行手段による上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得手段により取得した上記立体構造データとを比較する予測結果比較手段と、

上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、

上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段と、

を備えたことを特徴とする相互作用部位予測装置。

【請求項 3】 上記断片構造予測プログラムの上記断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定手段、

をさらに備え、

上記フラストレーション計算手段は、上記確信度情報設定手段により設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする請求項 1 または 2 に記載の相互作用部位予測装置。

【請求項 4】 目的のタンパク質の一次配列情報を入力する入力ステップと

タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、

上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果を比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法。

【請求項 5】 目的のタンパク質の一次配列情報を入力する入力ステップと

上記目的のタンパク質の立体構造データを取得する立体構造データ取得ステッ

ブと、

タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、

上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得ステップにより取得した上記立体構造データとを比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法。

【請求項 6】 上記断片構造予測プログラムの上記断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定ステップ、

をさらに含み、

上記フラストレーション計算ステップは、上記確信度情報設定ステップにより設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする請求項 4 または 5 に記載の相互作用部位予測方法。

【請求項 7】 目的のタンパク質の一次配列情報を入力する入力ステップと

タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、

上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラム

の断片構造予測結果を比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法をコンピュータに実行させるためのプログラム。

【請求項 8】 目的のタンパク質の一次配列情報を入力する入力ステップと

上記目的のタンパク質の立体構造データを取得する立体構造データ取得ステップと、

タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、

上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得ステップにより取得した上記立体構造データとを比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法をコンピュータに実行させるためのプログラム。

【請求項 9】 上記断片構造予測プログラムの上記断片構造予測結果に対す

る確信度を示す確信度情報を設定する確信度情報設定ステップ、
をさらに含み、

上記フラストレーション計算ステップは、上記確信度情報設定ステップにより
設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記
フラストレーションを計算することを特徴とする請求項 7 または 8 に記載のプロ
グラム。

【請求項 1 0】 上記請求項 7 ～ 9 のいずれか一つに記載されたプログラム
を記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、相互作用部位予測装置、相互作用部位予測方法、プログラム、およ
び、記録媒体に関し、特に、局所部位のフラストレーションに基づいて相互作用
部位を予測する相互作用部位予測装置、相互作用部位予測方法、プログラム、お
よび、記録媒体に関する。

【 0 0 0 2 】

【従来の技術】

タンパク質が働くため、すなわちタンパク質が一定の機能を表すためには他の
タンパク質や基質等と何らかの相互作用をすることが必要である。したがって、
タンパク質の相互作用部位を決定することは、創薬等の分野においては極めて重
要な研究テーマであり、従来より、バイオインフォマティクス分野などにおいて
、タンパク質の一次配列情報（アミノ酸配列情報）に対してモチーフ検索を実行
するなどの手法によりタンパク質の相互作用部位を解析する技術が開発されてい
る。すなわち、既知の相互作用部位に特異的に存在するアミノ酸配列を探索する
ことにより、タンパク質の相互作用部位を予測している。

【 0 0 0 3 】

【発明が解決しようとする課題】

しかしながら、従来のモチーフ検索などによる相互作用部位の解析手法は、既
知の相互作用部位については解析することができるが、未知の相互作用部位につ

いては解析することができないというシステム構造上の基本的問題点を有していた。以下、この問題点の内容について、一層具体的に説明する。

【0004】

すなわち、従来の相互作用部位の解析手法は、予め判明している相互作用部位に特異的な一次配列をモチーフデータベースなどに登録し、その情報を用いて相互作用部位の予測を行うものであるため、今までに発見されていない相互作用部位については、その解析をすることができなかった。従って、今まで発見されていない未知の相互作用部位をバイオインフォマティクス技術を用いてコンピュータ上で予測する際には、まったく別の手法を用いる必要があるが、有効な手法は未だ確立していない。

【0005】

ところで、タンパク質のネイティブの立体構造は、できるだけアミノ酸間の相互作用にフラストレーションが無いような構造に形作られている。すなわち、タンパク質のエネルギー曲面はフラストレーションが無いような全体構造（ネイティブ構造）へ向かって漏斗状に設計されていると言われている（*folding funnel*）。ここで、ネイティブ構造はフラストレーションが少ない構造ではあるが、要素間に働く相互作用の複雑さ、自由度の大きさ、進化的な経緯などの点から、フラストレーションが完全に無くなっているわけではない。

【0006】

最近の計算機実験では、進化の産物であるタンパク質の漏斗状のエネルギー面は本質的に等方的ではなく、フラストレーションの大きい方向と小さな方向が存在している（異方性を持つ）ことがわかってきた（*anisotropic funnel*）。これは構造的に言うならば、局所的な構造にはフラストレーションの大きい構造と小さい構造があることを示している。そして、フラストレーションの大きい局所構造部位は、全体構造の安定化のために犠牲になっている構造部位であり、この部分は、全体構造の安定化のために歪んだ構造を取らざるを得ない状況下に置かれており、いわば全体構造の中における不安定部位であることを示している。

【0007】

一方、タンパク質の相互作用は二つの安定な全体構造をもつタンパク質同士が作用することにより、さらに安定化する過程であるといえる。タンパク質が相互作用する際の構造の変化について更に説明すると、タンパク質AとBとが相互作用する際には、タンパク質Aの一部分の構造とタンパク質Bの一部分の構造に変化が起こり結合する。

【0008】

変化が起こる一部分の構造として考えられる部位について考察すると、まず、局所的に見ても、かつ全体的に見ても安定な局所構造は、それ以上安定化する必要性がない。一方、全体的に見れば安定であるが局所的に見ると不安定である部分は、他のタンパク質等と結合することによりその部分が安定化し、さらに結合することにより全体もさらに安定化するという状況が考えられる。すなわち、局所的に不安定な構造領域は、タンパク質相互作用部位である可能性が比較的高いと考えることができる。このように局所的に不安定な部位を一次配列から予測することにより、相互作用部位の候補を挙げることができるようになる可能性がある。

【0009】

タンパク質の立体構造は、配列情報から一意に決まると言われている。これは、配列空間と構造空間との間には何らかの相関があるということを意味する。ここで、配列空間と構造空間（ネイティブ構造の空間）の大きさを比較してみると、配列空間の方が大きいと言える。なぜなら、進化的に見ると、配列が多少変化したとしても構造はあまり変化しているようには見えないからである。このことを換言すると、構造の方が配列よりも進化的に保守性が強いとも言える。

【0010】

また、近年の進化的に良く似たタンパク質の構造解析によって、似たような配列をもつタンパク質は、似たような全体構造をもつということが明らかになってきた。全体は部分の組み合わせから作られていると考え、タンパク質の全体構造に対して成り立つであろう上記のような経験則が、タンパク質の一部を切り出した部分に関してもある程度成り立つのではないかと考えることができる。

【0011】

実際に、タンパク質の似たような部分配列は似たような部分構造となるという部分配列と構造とに相関のあるものが存在している。最近の研究では、部分配列と部分構造との相関を利用して、部分配列から全体構造を組み立てるという試みが試されてきている。

【0012】

例えば、Kim T. Simons等「Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions」J. Mol. Biol. (1997) 268, 209p-225p (以下「文献1」という) や、Christopher Bystroff等「Prediction of Local Structure in Proteins Using a Library of Sequence-Structure Motifs」J. Mol. Biol. (1998) 281, 565p-577p (以下「文献2」という) により開示された研究においては、部分配列に対応する構造をクラスタリングすることにより、広大な構造 (folding) 空間を狭めることが可能となり、フォールディング・シミュレーションの計算時間の短縮にも繋がっている。

【0013】

ここで、文献1には、localな配列によってlocal構造は特定の偏った構造に制限されるので構造空間は小さくなること、その構造は配列上近いタンパク質の構造に類似していること、配列のプロファイルをマルチプルアラインメントにより求め、問合せ (query) 配列との距離の近さを求めることなどが開示されている。

【0014】

また、文献2には、断片構造と配列の相関がある場合には、断片の配列傾向から限られた数の構造の候補がとりだせること、構造に関しては二つの構造指標を用いてクラスタリングを行い、また、配列についてはfrequency profileの距離を使って計算し、クラスター作成は配列が似ているものの中から

ら構造が似ているものを探してクラスター化することにより実際に配列-構造断片のクラスターを作成することが開示されている。

【0015】

ここで、部分配列-部分構造クラスターは、配列と相関の高い部分構造予測を行うためのものであるが、相関の強さ（似た配列がその構造である確信度）を用いて、全体構造の中で（配列と構造との間の相関の観点から）局所的に安定な部位（相関の強い部位）や、不安定な部位を特定することが可能であると考えられる。このクラスターはまた、データセット、部分配列の長さ、クラスターの大きさなどの観点から様々な分類が可能である。

【0016】

部分配列と部分構造との間の相関の強い部分は、その部分配列のみで全体構造におけるその部位の構造が決まっている部位（つまり、全体構造の中でも安定している部位）である可能性が高いといえる。一方、部分配列と部分構造との間の相関の弱い部分は、その部分配列のみでは構造が決まらない（全体構造の中でその部分構造が決まる）部位である可能性が高い。

【0017】

よって、フラストレーションが大きい（不安定な）局所部位の候補としては、部分配列と部分構造の相関が弱い部分や、様々なクラスターでの結果において、確信度の高い結果がまちまちである部分や、フォールディング・シミュレーション（*folding simulation*）を実行した後、確信度が高く予想された初期構造とは異なる構造を取っている部分や、まわりの局所構造と整合性を欠く部分が考えられる。

【0018】

このようなフラストレーションの大きい局所部位の断片構造予測においては、上述した様々な手法における処理の仕方の差が大きいと言えるだろう。すなわち、種々の手法による誤差が大きくなる、すなわち精度の悪くなる部位等が、フラストレーションの大きい局所部位である可能性が高いといえる。よって、様々な手法における断片構造予測の結果比較を通して、フラストレーションの比較的大きな局所部位の予測が可能になると考えることができる。

【 0 0 1 9 】

また、タンパク質の立体構造データが分かっている場合、すなわち、既存の PDB などに立体構造データが登録されているタンパク質は、その全体構造が既知であるため、様々な断片構造予測手法の予測結果とその実際の構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になると考えられる。

【 0 0 2 0 】

本発明は上記問題点に鑑みてなされたもので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することのできる、相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することを目的としている。

【 0 0 2 1 】

【課題を解決するための手段】

このような目的を達成するため、請求項 1 に記載の相互作用部位予測装置は、目的のタンパク質の一次配列情報を入力する入力手段と、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行手段と、上記断片構造予測プログラム実行手段による上記断片構造予測プログラムの断片構造予測結果を比較する予測結果比較手段と、上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段とを備えたことを特徴とする。

【 0 0 2 2 】

この装置によれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果を比較し、比較結果に基づいて、目

的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる。

【 0 0 2 3 】

また、請求項 2 に記載の相互作用部位予測装置は、目的のタンパク質の一次配列情報を入力する入力手段と、上記目的のタンパク質の立体構造データを取得する立体構造データ取得手段と、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行手段と、上記断片構造予測プログラム実行手段による上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得手段により取得した上記立体構造データとを比較する予測結果比較手段と、上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段とを備えたことを特徴とする。

【 0 0 2 4 】

この装置によれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の立体構造データを取得し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果と、取得した立体構造データとを比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、断片構造予測プログラムの予測結果と目的のタンパク質の実際の断片構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。

【 0 0 2 5 】

また、請求項 3 に記載の相互作用部位予測装置は、請求項 1 または 2 に記載の相互作用部位予測装置において、上記断片構造予測プログラムの上記断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定手段をさらに備え、上記フラストレーション計算手段は、上記確信度情報設定手段により設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする。

【 0 0 2 6 】

これはフラストレーション計算の一例を一層具体的に示すものである。この装置によれば、断片構造予測プログラムの断片構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる断片構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができるようになる。

【 0 0 2 7 】

また、本発明は相互作用部位予測方法に関するものであり、請求項 4 に記載の相互作用部位予測方法は、目的のタンパク質の一次配列情報を入力する入力ステップと、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果を比較する予測結果比較ステップと、上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップとを含むことを特徴とする。

【 0 0 2 8 】

この方法によれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果を比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる。

【0029】

また、請求項5に記載の相互作用部位予測方法は、目的のタンパク質の一次配列情報を入力する入力ステップと、上記目的のタンパク質の立体構造データを取得する立体構造データ取得ステップと、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得ステップにより取得した上記立体構造データとを比較する予測結果比較ステップと、上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップとを含むことを特徴とする。

【0030】

この方法によれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の立体構造データを取得し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果と、取得した立体構造データとを比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算さ

れた局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、断片構造予測プログラムの予測結果と目的のタンパク質の実際の断片構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。

【 0 0 3 1 】

また、請求項 6 に記載の相互作用部位予測方法は、請求項 4 または 5 に記載の相互作用部位予測方法において、上記断片構造予測プログラムの上記断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定ステップをさらに含み、上記フラストレーション計算ステップは、上記確信度情報設定ステップにより設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする。

【 0 0 3 2 】

これはフラストレーション計算の一例を一層具体的に示すものである。この方法によれば、断片構造予測プログラムの断片構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる断片構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができるようになる。

【 0 0 3 3 】

また、本発明はプログラムに関するものであり、請求項 7 に記載のプログラムは、目的のタンパク質の一次配列情報を入力する入力ステップと、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果を比較する予測結果比較ステップと、上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、上記フラストレー

ション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップとを含む相互作用部位予測方法をコンピュータに実行させることを特徴とする。

【0034】

このプログラムによれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果を比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる。

【0035】

また、請求項8に記載のプログラムは、目的のタンパク質の一次配列情報を入力する入力ステップと、上記目的のタンパク質の立体構造データを取得する立体構造データ取得ステップと、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行ステップと、上記断片構造予測プログラム実行ステップによる上記断片構造予測プログラムの断片構造予測結果と、上記立体構造データ取得ステップにより取得した上記立体構造データとを比較する予測結果比較ステップと、上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップとを含む相互作用部位予測方法をコンピュータに実行させることを特徴とする。

【0036】

このプログラムによれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の立体構造データを取得し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果と、取得した立体構造データとを比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、断片構造予測プログラムの予測結果と目的のタンパク質の実際の断片構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。

【 0 0 3 7 】

また、請求項 9 に記載のプログラムは、請求項 7 または 8 に記載のプログラムにおいて、上記断片構造予測プログラムの上記断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定ステップをさらに含み、上記フラストレーション計算ステップは、上記確信度情報設定ステップにより設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする。

【 0 0 3 8 】

これはフラストレーション計算の一例を一層具体的に示すものである。このプログラムによれば、断片構造予測プログラムの断片構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる断片構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができるようになる。

【 0 0 3 9 】

また、本発明は記録媒体に関するものであり、請求項 1 0 に記載の記録媒体は、上記請求項 7 ～ 9 のいずれか一つに記載されたプログラムを記録したことを特

徴とする。

【0040】

この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、請求項7～9のいずれか一つに記載されたプログラムをコンピュータを利用して実現することができ、これら各プログラムと同様の効果を得ることができる。

【0041】

【発明の実施の形態】

以下に、本発明にかかる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

【0042】

〔本発明の概要〕

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図1は本発明の基本原理を示す原理構成図である。

【0043】

本発明は、概略的に、以下の基本的特徴を有する。すなわち、本発明は、断片構造クラスターを用いて相互作用部位を予測すること、断片構造の予測結果から相互作用部位を予測すること、様々な断片構造予測の予測結果の差異から相互作用予測部位を予測すること、断片構造の予測結果の差異が表れる部位が全体構造と部分構造におけるフラストレーションの高いストレス部位であると予測すること、構造的なストレス部位（比較的構造的に不安定な部分）は他と相互作用する部位である可能性があると予測すること等を特徴とする。以下に、本発明の基本原理を説明する。

【0044】

まず、利用者は、目的のタンパク質の一次配列情報である目的配列データ10を本発明の相互作用部位予測装置に入力する。この目的配列データ10の入力は、例えば、利用者がSWISS-PROTやPIRやTrEMBL等の外部のデータベースに登録された一次配列情報を選択することにより入力してもよく、ま

た、所望の一次配列情報を直接入力してもよい。

【 0 0 4 5 】

ついで、本発明の相互作用部位予測装置は、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラム 2 0 a ~ 2 0 d に対して入力された目的配列データ 1 0 の断片構造予測シミュレーションを実行する。ここで、断片構造予測プログラム 2 0 a ~ 2 0 d は、例えば、文献 1 または文献 2 に開示された手法や、スレッディング法や、アブイニシオ法などを利用して断片構造予測シミュレーションを実行する。

【 0 0 4 6 】

ついで、本発明の相互作用部位予測装置は、各断片構造予測プログラム 2 0 a ~ 2 0 d の断片構造予測結果 3 0 a ~ 3 0 d を比較する (6 0) 。すなわち、目的配列データ 1 0 に対応する各予測プログラムの実行結果を並置して比較する (3 0 a ~ 3 0 d) 。

【 0 0 4 7 】

そして、本発明の相互作用部位予測装置は、この比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算する (7 0) 。すなわち、比較結果から各予測結果データ (3 0 a ~ 3 0 b) で異なる断片構造を予測した局所部分を抽出して、その部分のフラストレーションを計算する。既存の断片構造予測プログラム 2 0 a ~ 2 0 d は、基本的には一次配列情報のうち一部の局所配列を見て予測を行うが、断片構造はタンパク質の全体の構造との関係で最終的に決定されるので、全体と局所部分との間に整合性が取れない部位、すなわちフラストレーションが大きい局所的部位では断片構造予測結果は外れる場合が多い。従って、複数のプログラムにおいて予測結果が異なる局所部分については、フラストレーションが大きいと仮定することができる。

【 0 0 4 8 】

フラストレーションの計算方法は、例えば、異なる予測結果データを出力した断片構造予測プログラム数に応じてフラストレーションを増減してもよく、あるいは、異なる予測結果のそれぞれの構造における確信度の平均値や分散値などに応じてフラストレーションを増減してもよく、また、その部分のアミノ酸配列を

分子力学または分子動力学などの手法を用いてエネルギー量を求めそのエネルギー量を用いてフラストレーションを計算してもよい。

【 0 0 4 9 】

そして、本発明の相互作用部位予測装置は、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する（80）。すなわち、例えば、一定の閾値を超えるフラストレーションが存在する局所部分（61）を相互作用部位と予測する。

【 0 0 5 0 】

また、本発明の相互作用部位予測装置は、例えば、PDBやSCOPなどの外部のデータベースに目的のタンパク質の立体構造データが登録されている場合にはその立体構造データ40を取得し、予測結果の比較において用いる。すなわち、目的のタンパク質が実際に採る立体構造データ40と、予測プログラムの予測結果データ30a～30dとを比較する。

【 0 0 5 1 】

そして、実際の立体構造データ40と予測プログラムの予測結果データ30a～30dとが異なっている部分については、高いフラストレーションを計算する。タンパク質の立体構造データが分かっている場合、すなわち、既存のPDBなどに立体構造データが登録されているタンパク質は、その全体構造が既知であるため、様々な断片構造予測手法の予測結果とその実際の構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。例えば、実際の立体構造データ40と異なる予測結果データを出力した断片構造予測プログラム数に応じてフラストレーションを増減してもよい。

【 0 0 5 2 】

さらに、本発明の相互作用部位予測装置は、断片構造予測プログラム20a～20dの断片構造予測結果データ30a～30dに対する確信度を示す確信度情報50を設定する。すなわち、断片構造予測プログラム20a～20dのシミュレーション精度を実際の立体構造データなどに基づいて設定する。

【 0 0 5 3 】

そして、本発明の相互作用部位予測装置は、設定された確信度情報および比較結果に基づいて、局所部分の上記フラストレーションを計算する。すなわち、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる断片構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができる。

【 0 0 5 4 】

すなわち、本発明は、それぞれの手法における構造予測の結果とその構造への確信度を解析し、部分配列→部分構造の相関が弱い部分や、様々なクラスターでの結果において確信度の高い結果がまちまちである部分や、フォールディング・シミュレーション（*folding simulation*）を実行した後、確信度が高く予想された初期構造とは異なる構造を取っている部分や、まわりの局所構造と整合性を欠く部分等に対して高いフラストレーションとなるように候補部位に関する得点を計算し、計算した結果をもとに相互作用部位として可能性の高い部位を順に得点をつけて抽出することができる。

【 0 0 5 5 】

〔システム構成〕

まず、本システムの構成について説明する。図2は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に係る部分のみを概念的に示している。本システムは、概略的に、相互作用部位予測装置100と、配列情報や立体構造等に関する外部データベースおよびホモロジー検索や断片構造予測等の外部プログラム等を提供する外部システム200とを、ネットワーク300を介して通信可能に接続して構成されている。

【 0 0 5 6 】

図2においてネットワーク300は、相互作用部位予測装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネット等である。

【 0 0 5 7 】

図2において外部システム200は、ネットワーク300を介して、相互作用部位予測装置100と相互に接続され、利用者に対して配列情報や立体構造等に

関する外部データベース、および、ホモロジー検索やモチーフ検索や断片構造予測等の外部プログラムを実行するウェブサイトを提供する機能を有する。

【 0 0 5 8 】

ここで、外部システム 2 0 0 は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム 2 0 0 の各機能は、外部システム 2 0 0 のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

【 0 0 5 9 】

図 2 において相互作用部位予測装置 1 0 0 は、概略的に、相互作用部位予測装置 1 0 0 の全体を統括的に制御するCPU等の制御部 1 0 2、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部 1 0 4、入力装置 1 1 2 および出力装置 1 1 4 に接続される入出力制御インターフェース部 1 0 8、および、各種のデータベースやテーブル（予測結果データベース 1 0 6 a ～タンパク質構造データベース 1 0 6 c）を格納する記憶部 1 0 6 を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。さらに、この相互作用部位予測装置 1 0 0 は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク 3 0 0 に通信可能に接続されている。

【 0 0 6 0 】

図 2 において記憶部 1 0 6 に格納される各種のデータベースやテーブル（予測結果データベース 1 0 6 a ～タンパク質構造データベース 1 0 6 c）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

【 0 0 6 1 】

これら記憶部 1 0 6 の各構成要素のうち、予測結果データベース 1 0 6 a は、断片構造予測プログラムの予測結果などに関する情報を格納する予測結果情報格納手段である。図 3 は、予測結果データベース 1 0 6 a に格納される情報の一例

を示す図である。

【 0 0 6 2 】

この予測結果データベース 1 0 6 a に格納される情報は、図 3 に示すように、目的のタンパク質の一次配列情報（アミノ酸配列情報）である目的配列データ、タンパク質構造データベースから取得した目的配列データの立体構造データ、および、各断片構造予測プログラムの予測結果データを相互に関連付けて構成されている。

【 0 0 6 3 】

また、確信度情報データベース 1 0 6 b は、断片構造予測プログラムの断片構造予測結果データに対する確信度を示す確信度情報を格納する予測結果情報格納手段である。例えば、シミュレーション結果の精度の標準的な値（例えば、断片構造予測結果と実際の立体構造データとの一致率であるシミュレーション精度が 6 0 % の場合など）の確信度を 1 とし、標準値より精度が高い場合にはその精度に応じて確信度の値を大きくし、また、標準値より精度が低い場合にはその精度に応じて確信度の値を小さくしてもよい。さらに、断片構造予測プログラムごと、構造ごと、各配列におけるアミノ酸ごとに確信度を設定してもよい。すなわち、例えば、ある断片構造予測プログラムがある配列のあるアミノ酸についてその断片構造を予測するとき、その構造が a 構造である確信度、b 構造である確信度等をそれぞれ別々に設定してもよい。

【 0 0 6 4 】

また、タンパク質構造データベース 1 0 6 c は、タンパク質の立体構造データを格納したデータベースである。タンパク質構造データベース 1 0 6 c は、インターネットを経由してアクセスする外部のタンパク質構造データベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの配列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。

【 0 0 6 5 】

また、図 2 において、通信制御インターフェース部 1 0 4 は、相互作用部位予測装置 1 0 0 とネットワーク 3 0 0 （またはルータ等の通信装置）との間におけ

る通信制御を行う。すなわち、通信制御インターフェース部 1 0 4 は、他の端末と通信回線を介してデータを通信する機能を有する。

【 0 0 6 6 】

また、図 2 において、入出力制御インターフェース部 1 0 8 は、入力装置 1 1 2 や出力装置 1 1 4 の制御を行う。ここで、出力装置 1 1 4 としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置をモニタとして記載する）。また、入力装置 1 1 2 としては、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

【 0 0 6 7 】

また、図 2 において、制御部 1 0 2 は、OS (Operating System) 等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 1 0 2 は、機能概念的に、目的配列入力部 1 0 2 a、断片構造予測プログラム実行部 1 0 2 b、断片構造予測プログラム 1 0 2 c、予測結果比較部 1 0 2 d、フラストレーション計算部 1 0 2 e、相互作用部位予測部 1 0 2 f、立体構造データ取得部 1 0 2 g、および、確信度情報設定部 1 0 2 h を備えて構成されている。

【 0 0 6 8 】

このうち、目的配列入力部 1 0 2 a は、目的のタンパク質の一次配列情報（目的配列データ）を入力する入力手段である。また、断片構造予測プログラム実行部 1 0 2 b は、断片構造予測プログラムに対して入力手段により入力された一次配列情報（目的配列データ）の断片構造予測シミュレーションを実行させる断片構造予測プログラム実行手段である。また、断片構造予測プログラム 1 0 2 c は、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムである。

【 0 0 6 9 】

また、予測結果比較部 1 0 2 d は、断片構造予測プログラムの断片構造予測結果を比較する予測結果比較手段、および、断片構造予測プログラムの断片構造予

測結果と立体構造データ取得手段により取得した立体構造データとを比較する予測結果比較手段である。また、フラストレーション計算部 1 0 2 e は、予測結果比較手段による比較結果に基づいて、目的のタンパク質の一次配列情報（目的配列データ）の局所部分のフラストレーションを計算するフラストレーション計算手段、および、確信度情報設定手段により設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するフラストレーション計算手段である。

【 0 0 7 0 】

また、相互作用部位予測部 1 0 2 f は、フラストレーション計算手段により計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する相互作用部位予測手段である。また、立体構造データ取得部 1 0 2 g は、目的のタンパク質の立体構造データを取得する立体構造データ取得手段である。また、確信度情報設定部 1 0 2 h は、断片構造予測プログラムの断片構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定手段である。なお、これら各部によって行なわれる処理の詳細については、後述する。

【 0 0 7 1 】

〔システムの処理〕

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 4 ～図 7 を参照して詳細に説明する。

【 0 0 7 2 】

〔メイン処理〕

次に、メイン処理の詳細について図 4 を参照して説明する。図 4 は、本実施の形態における本システムのメイン処理の一例を示すフローチャートである。

【 0 0 7 3 】

まず、相互作用部位予測装置 1 0 0 は、目的配列入力部 1 0 2 a の処理により、利用者に対して目的のタンパク質の一次配列情報（目的配列データ）を入力させる（ステップ S A - 1 ）。

【 0 0 7 4 】

ついで、相互作用部位予測装置 1 0 0 は、立体構造データ取得部 1 0 2 g の処

理により、利用者が入力した目的配列データの立体構造データを取得する（ステップSA-2）。

【0075】

ここで、ステップSA-2において、立体構造データ取得部102gにより実行される立体構造データ取得処理の詳細について図5を参照して説明する。図5は、本実施の形態における本システムの立体構造データ取得処理の一例を示すフローチャートである。

【0076】

まず、立体構造データ取得部102gは、タンパク質構造データベース106cを参照して目的配列データが登録されているか判定する（ステップSB-1）。ステップSB-1において、目的配列データがタンパク質構造データベース106cに登録されている場合には、立体構造データ取得部102gは、タンパク質構造データベース106cから目的配列データの立体構造データを取得して予測結果データベース106aの所定の記憶領域に格納する（ステップSB-2）。

【0077】

一方、ステップSB-1において、目的配列データがタンパク質構造データベース106cに登録されていない場合には、立体構造データ取得部102gは、目的配列データに類似する配列のタンパク質の立体構造データがタンパク質構造データベース106cに存在するか判断する（ステップSB-3）。すなわち、立体構造データ取得部102gは、例えば、配列間のホモロジーを判定するプログラムを用いて、目的配列データと、タンパク質構造データベース106cに登録された構造既知のタンパク質に対応する配列データとを比較して、高い相同性を有する配列データ（目的配列データの一部に対応するものであってもよい）が存在するか否かを判断する。

【0078】

ステップSB-3において、目的配列データに類似する配列のタンパク質の立体構造データがタンパク質構造データベース106cに存在する場合には、立体構造データ取得部102gは、類似部分の立体構造データを予測結果データベ-

ス106aの所定の記憶領域に格納する（ステップSB-4）。なお、目的配列データのうち一部分について立体構造データが存在する場合には、立体構造データが存在する部分について、当該立体構造データを予測結果データベース106aに格納する。

【0079】

一方、ステップSB-3において、目的配列データに類似する配列のタンパク質の立体構造データがタンパク質構造データベース106cに存在しない場合には、立体構造データ取得処理が終了する。

【0080】

再び図4に戻り、相互作用部位予測装置100は、断片構造予測プログラム実行部102bの処理により、目的配列データを1つまたは2つ以上の断片構造予測プログラム102cに実行させる（ステップSA-3）。すなわち、断片構造予測プログラム実行部102bは、例えば、目的配列データを所定のフォーマットに変換したり、また、目的配列データに所定のヘッダ情報などを追加したりすることにより、各断片構造予測プログラム102cの入力形式を合わせた後、断片構造予測プログラム102cを実行する。ここで、断片構造予測プログラム102cは、相互作用部位予測装置100の内部に存在するプログラムであってもよく、また、ネットワーク300を介してリモートで実行することができる外部システム200の外部プログラムであってもよい。

【0081】

ここで、断片構造予測プログラム実行部102bは、以下のいずれかの手法により断片構造を求めてもよい。

【0082】

(1) オーソログ (ortholog) 相同性解析

ある遺伝子X、YのオーソログX'、Y'があり、X'、Y'が相互作用することが既知の場合、X、Yも相互作用する可能性が高いだろうと考える解析手法である。

【0083】

(2) ロゼッタストーン (Rosetta Stone) 法

ある生物種の遺伝子X、Yが、他生物種遺伝子の遺伝子Zと、例えば、前半部分はXと似ている、後半部分はYと似ているという状況のとき、進化により以前は一つの遺伝子ZだったものがX、Yに分離したと考え、XとYは相互作用する可能性が高いと考える解析手法である。

【0084】

(3) スレッディング (threading) 法

構造情報から3Dプロファイルを作り、これを用いて配列とアラインメントを行うことにより、構造予測を行う手法である。

【0085】

(4) トポロジー・フィンガープリント法 (topology fingerprint 法)

あるタンパク質Aの構造の内外性、距離マップなどを用いて、その構造の topology print と呼ばれるパラメータ (その構造におけるエネルギー関数の係数等に対応する) を抽出し、このパラメータをタンパク質Bの配列に適用することにより、アラインメントを行う手法である。

【0086】

(5) モチーフ検索法

モチーフを登録したデータベース (例えば、PROSITE、Pfamなど) に登録されているモチーフをキーにして検索することにより機能 (構造) を予測する手法である。同じ機能モチーフを持っているタンパク質は同様の機能 (構造) をもつ可能性が高いと考える。

【0087】

(6) モジュール検索 (3Dキーノート) 法

モジュールとは、球状タンパク質においてコンパクトな構造をとる15残基程度のアミノ酸配列であり、イントロンがモジュールの境界とよく対応していると言われていることから、色々なタンパク質に存在する共通のモジュール構造からアミノ酸配列のパターンを抽出し、それをキーにして検索することにより構造予測を行う手法である。機能に関わるモジュール構造から抽出された特徴的なアミノ酸配列の類似性は、それらのタンパク質間での機能の類似性を意味すると考え

られる。

【0088】

(7) 進化追跡法 (Evolutionary trace method)

あるタンパク質の立体構造とそのオーソログのマルチプルアラインメントが与えられたとき、配列の変化を進化的に追うことによって、基質結合部位や、別のタンパク質との相互作用部位を予測する方法である。

【0089】

(8) ホモロジープロファイル法 (HMM法)

ある配列Aがあったとき、そのファミリーやオーソログなどの関連タンパク質があり、それらのアラインメントがとれた場合に、配列Aまたはそのファミリーなどに関してプロファイル行列を作成する。これと配列B（あるいは配列Bに関して同様なことを行った配列Bのプロファイル）とのアラインメントをとる手法である。本手法によれば、配列Aと配列Bとの間のアラインメントよりも、より遠くの関係を検索することができる。

【0090】

(9) Fugue

類縁タンパク質のアミノ酸配列保存性を、立体構造上どの部分で保存されているかにより分類し、この情報を用いて従来のものより感度の高いBLOSUMのような構造-配列の置換マトリックスを作成する手法である。本手法によれば、マトリックスを用いてアラインメントすることにより、より感度の高いアラインメントが可能となる。

【0091】

なお、本発明は、これらの手法に限定されず、断片構造を予測するいかなる手法を用いてもよい。

【0092】

ついで、断片構造予測プログラム実行部102bは、各断片構造予測プログラム102cのシミュレーション結果である断片構造予測結果を予測結果データベース106aの所定の記憶領域に格納する（ステップSA-4）。

【0093】

ついで、相互作用部位予測装置 1 0 0 は、予測結果比較部 1 0 2 d の処理により、予測結果データベース 1 0 6 a に格納された目的配列データに対する各断片構造予測プログラム 1 0 2 c の断片構造予測結果を比較する（ステップ S A - 5）。すなわち、予測結果比較部 1 0 2 d は、各断片構造予測プログラム 1 0 2 c の断片構造予測結果について、目的配列データの先頭から最後まで各予測結果を比較する。なお、ステップ S A - 2 において、断片構造予測プログラム実行部 1 0 2 b が目的配列データに対応する立体構造データを取得できた場合、すなわち、予測結果データベース 1 0 6 a に目的配列データの立体構造データが格納されている場合には、立体構造データと各断片構造予測プログラム 1 0 2 c の断片構造予測結果とを比較する。

【 0 0 9 4 】

ついで、相互作用部位予測装置 1 0 0 は、フラストレーション計算部 1 0 2 e の処理により、目的配列データの局所部分のフラストレーションのスコアを計算する（ステップ S A - 6）。ここで、図 6 は、本システムのフラストレーション計算部 1 0 2 e により実行されるフラストレーション実行処理の一例を示すフローチャートである。

【 0 0 9 5 】

図 6 に示すように、フラストレーション計算部 1 0 2 e によるフラストレーションのスコアの計算方法は、例えば、断片構造予測プログラムが異なる断片構造予測結果を出力した局所部分について、結果が異なる断片構造予測プログラムの数に応じてスコアを増減してもよく、あるいは、異なる予測結果のそれぞれの構造における確信度の平均値や分散値などに応じてフラストレーションを増減してもよく、また、断片構造予測プログラムが異なる断片構造予測結果を出力した局所部分について、アミノ酸配列を分子力学または分子動力学などの手法を用いてエネルギー量を求めそのエネルギー量を用いてフラストレーションを計算してもよい（ステップ S C - 1）。

【 0 0 9 6 】

また、フラストレーション計算部 1 0 2 e は、立体構造データと予測プログラムの断片構造予測結果とが異なっている部分について、高いフラストレーション

のスコアを計算してもよい（ステップSC-2）。例えば、立体構造データと異なる断片構造予測結果を出力した断片構造予測プログラムの数に応じてスコアを増減してもよい。

【0097】

また、フラストレーション計算部102eは、確信度情報データベース106bを参照して、予め確信度情報設定部102hの処理により格納された各断片構造予測プログラム102cの確信度情報を取得し、確信度情報に基づいてフラストレーションのスコアを計算してもよい（ステップSC-3）。すなわち、フラストレーション計算部102eは、シミュレーション精度の高い断片構造予測プログラム102cの断片構造予測結果に高い重みを付けて、フラストレーションのスコアを計算する。

【0098】

ここで、確信度情報設定部102hによる確信度情報の設定の一例を示す。まず、確信度情報設定部102hは、各断片構造予測プログラム102cの断片構造予測結果と、立体構造データとを比較して、各断片構造予測プログラム102cの断片構造予測結果の精度（一致率）を計算する。そして、確信度情報設定部102hは、各断片構造予測プログラム102cの精度の平均値を標準の確信度情報（例えば、1）として設定し、平均値以上の精度については標準の確信度情報よりも高い値（例えば、1より大きい数）を計算し、平均値以下の精度については標準の確信度情報よりも低い値（例えば、1より小さい数）を計算して、確信度情報データベース106bの所定の記憶領域に格納する。

【0099】

なお、確信度情報設定部102hは、各断片構造予測プログラム102cの確信度情報を各配列中の各アミノ酸（残基）毎に設定してもよい。すなわち、各断片構造予測プログラム102cによる配列の予測結果につき配列中のアミノ酸毎に断片構造予測プログラム102cの確信度情報を設定してもよい（例えば、配列中の一番目のアミノ酸について、プログラムAは、a構造の確信度情報が1.5であり、b構造の確信度情報が0.7であり、c構造の確信度情報が1.1である等）。

【0100】

また、確信度情報設定部102hは、各断片構造予測プログラム102cの確信度情報を構造毎に設定してもよい。すなわち、各断片構造予測プログラム102cによっては、特定の構造について高い精度を有するものや低い精度を有するものが存在するため、構造毎に断片構造予測プログラム102cの確信度情報を設定してもよい（例えば、プログラムAは、a構造の確信度情報が1.5であり、b構造の確信度情報が0.7であり、c構造の確信度情報が1.1である等）。

【0101】

再び図4に戻り、相互作用部位予測装置100は、相互作用部位予測部102fの処理により、計算された局所部分のフラストレーションのスコアにより目的配列データのうち相互作用部位となる局所部分を予測する（ステップSA-7）。すなわち、相互作用部位予測部102fは、例えば、一定の閾値を超えるフラストレーションのスコアである局所部分を相互作用部位と予測する。

【0102】

ついで、相互作用部位予測装置100は、配列データの相互作用部位の予測結果を出力装置114に出力する（ステップSA-8）。

【0103】

ここで、図7は、相互作用部位予測装置100の出力装置114に表示される相互作用部位予測結果の表示画面の一例を示す図である。この図に示すように相互作用部位予測結果の表示画面は、例えば、目的配列データの配列情報の表示領域MA-1、相互作用部位と予測される局所部分の表示領域MA-2およびMA-3、相互領域部位と予測される局所部分のフラストレーションのスコアの表示領域MA-4およびMA-5等を含んで構成されている。これにて、メイン処理が終了する。

【0104】

[他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内におい

て種々の異なる実施の形態にて実施されてよいものである。

例えば、相互作用部位予測装置 1 0 0 がスタンドアローンの形態で相互作用部位予測を行う場合を一例に説明したが、相互作用部位予測装置 1 0 0 とは別筐体で構成されるクライアント端末からの要求に応じて相互作用部位予測を行い、その予測結果を当該クライアント端末に返却するように構成してもよい。

【 0 1 0 5 】

また、断片構造予測プログラムが用いる予測手法は、いかなる方法を用いて予測するものであってもよい。

【 0 1 0 6 】

また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

【 0 1 0 7 】

この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

【 0 1 0 8 】

また、相互作用部位予測装置 1 0 0 に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

例えば、相互作用部位予測装置 1 0 0 の各サーバが備える処理機能、特に制御部にて行なわれる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該 CPU にて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じて相互作用部位予測装置 1 0 0 に機械的に読み取られる。

【 0 1 0 9 】

また、このプログラムは、相互作用部位予測装置 1 0 0 に対して任意のネット

ワークを介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

【0110】

また、記憶部106に格納される各種のデータベース等（予測結果データベース106a～タンパク質構造データベース106c）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

【0111】

また、相互作用部位予測装置100は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

【0112】

さらに、相互作用部位予測装置100の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部をCGI（Common Gateway Interface）を用いて実現してもよい。

【0113】

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信

する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

【0114】

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

【0115】

また、ネットワーク300は、相互作用部位予測装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VANや、パソコン通信網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網（アナログ／デジタルの双方を含む）や、CATV網や、IMT2000方式、GSM方式またはPDC／PDC-P方式等の携帯回線交換網／携帯パケット交換網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含んでもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

【0116】

【発明の効果】

以上詳細に説明したように、本発明によれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果を比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラスト

レーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

【 0 1 1 7 】

また、本発明によれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の立体構造データを取得し、タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラムに対して入力された一次配列情報の断片構造予測シミュレーションを実行させ、断片構造予測プログラムの断片構造予測結果と、取得した立体構造データとを比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、断片構造予測プログラムの予測結果と目的のタンパク質の実際の断片構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

【 0 1 1 8 】

さらに、本発明によれば、断片構造予測プログラムの断片構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる断片構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

【図面の簡単な説明】

【図 1】

本発明の基本原理を示す原理構成図である。

【図 2】

本発明が適用される本システムの構成の一例を示すブロック図である。

【図 3】

予測結果データベース 1 0 6 a に格納される情報の一例を示す図である。

【図 4】

本実施の形態における本システムのメイン処理の一例を示すフローチャートである。

【図 5】

本実施の形態における本システムの立体構造データ取得処理の一例を示すフローチャートである。

【図 6】

本システムのフラストレーション計算部 1 0 2 e により実行されるフラストレーション実行処理の一例を示すフローチャートである。

【図 7】

相互作用部位予測装置 1 0 0 の出力装置 1 1 4 に表示される相互作用部位予測結果の表示画面の一例を示す図である。

【符号の説明】

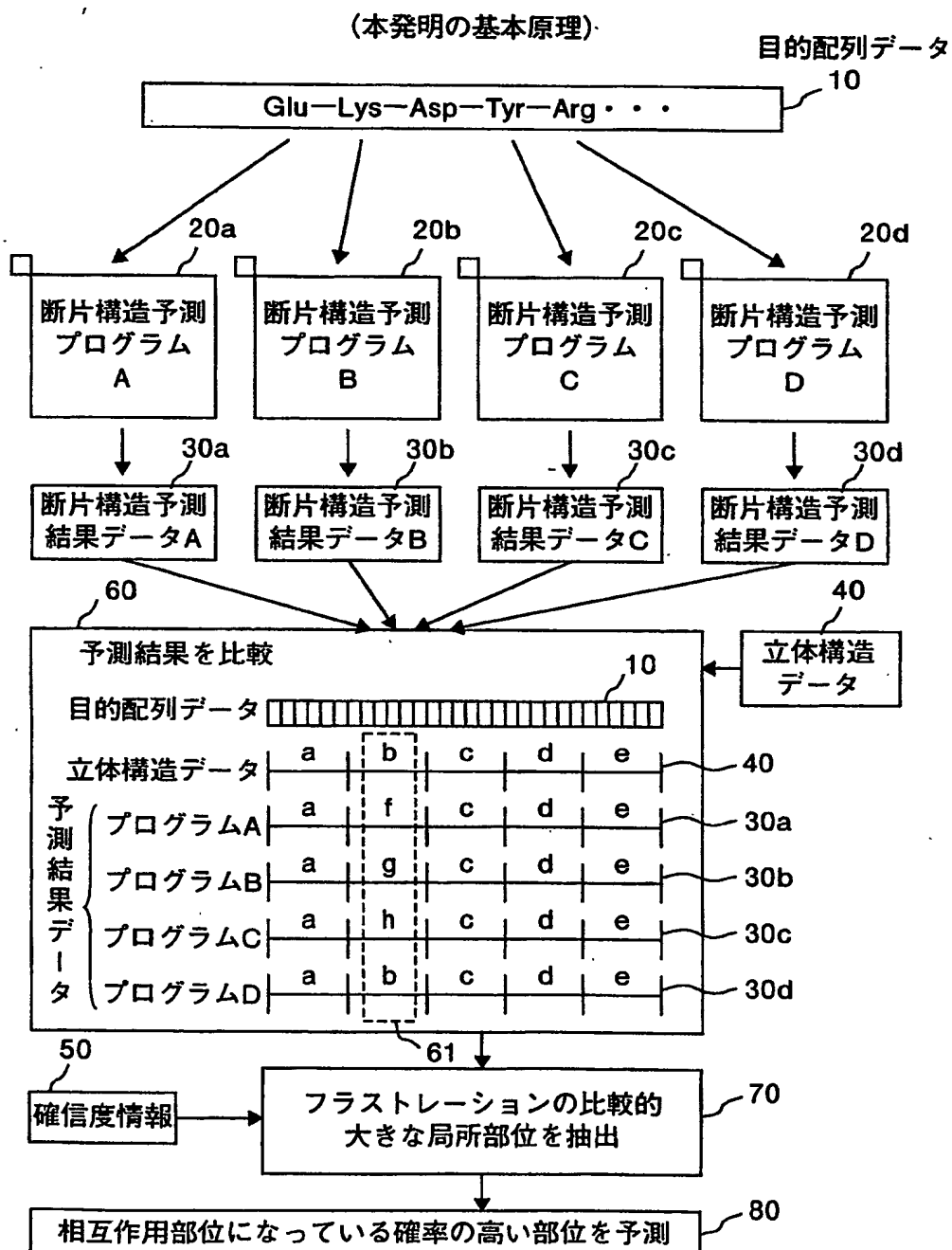
- 1 0 目的配列データ
- 2 0 断片構造予測プログラム
- 3 0 断片構造予測結果データ
- 4 0 立体構造データ
- 5 0 確信度情報
- 1 0 0 相互作用部位予測装置
 - 1 0 2 制御部
 - 1 0 2 a 目的配列入力部
 - 1 0 2 b 断片構造予測プログラム実行部
 - 1 0 2 c 断片構造予測プログラム
 - 1 0 2 d 予測結果比較部

- 102 e フラストレーション計算部
- 102 f 相互作用部位予測部
- 102 g 立体構造データ取得部
- 102 h 確信度情報設定部
- 104 通信制御インターフェース部
- 106 記憶部
 - 106 a 予測結果データベース
 - 106 b 確信度情報データベース
 - 106 c タンパク質構造データベース
- 108 入出力制御インターフェース部
- 112 入力装置
- 114 出力装置
- 200 外部システム
- 300 ネットワーク

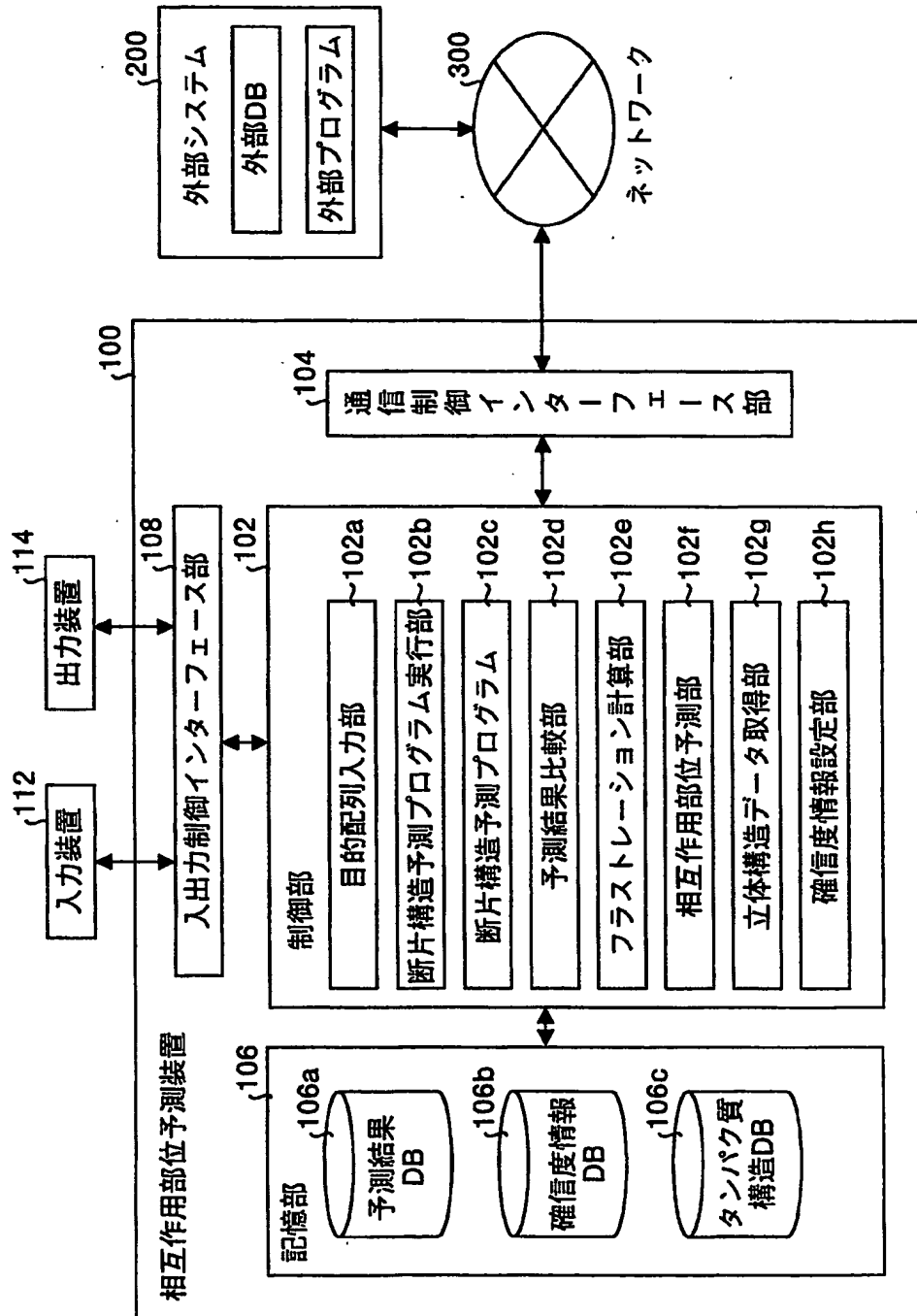
【書類名】

図面

【図 1】



【図 2】

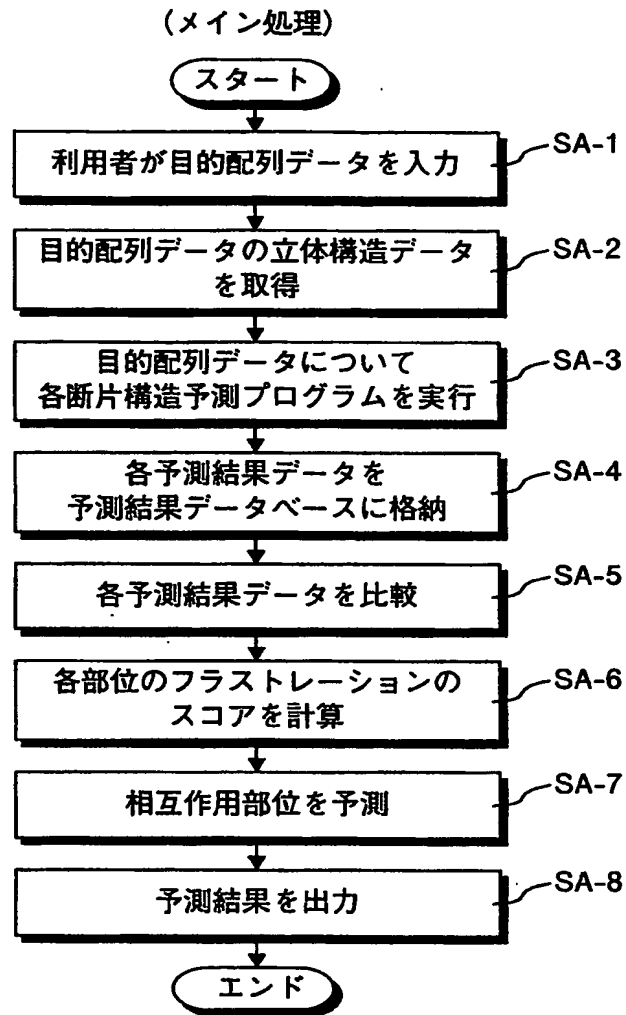


【図 3】

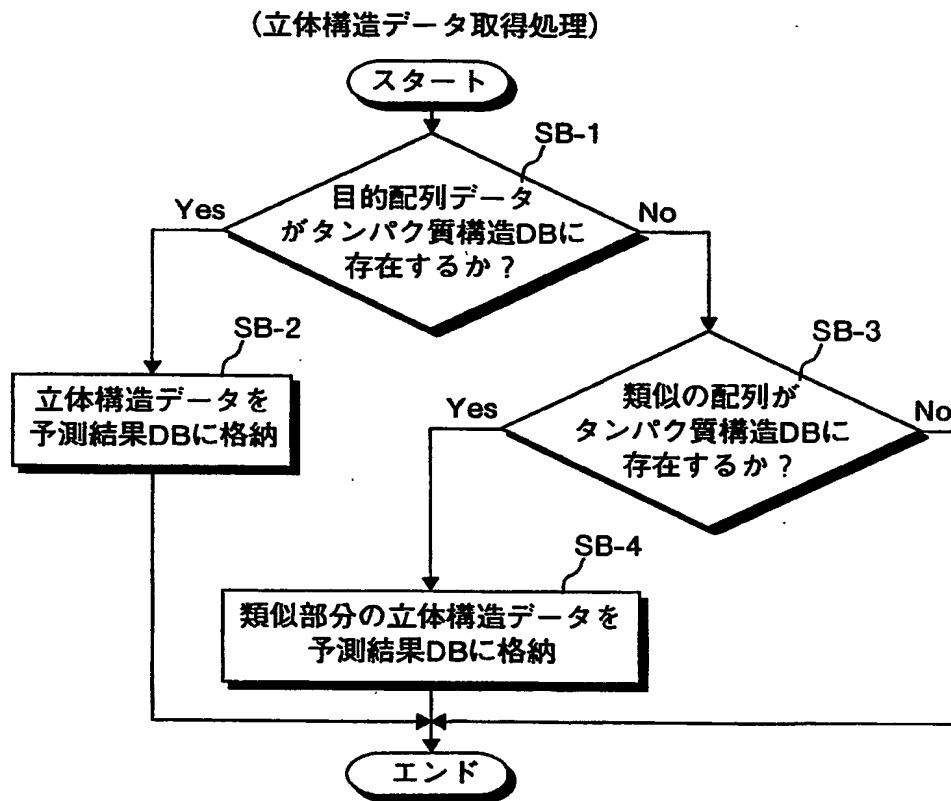
106a 予測結果データベース

目的配列 データ	立体構造 データ	予測結果データ				
		プログラムA	プログラムB	プログラムC	プログラムD	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

【図 4】

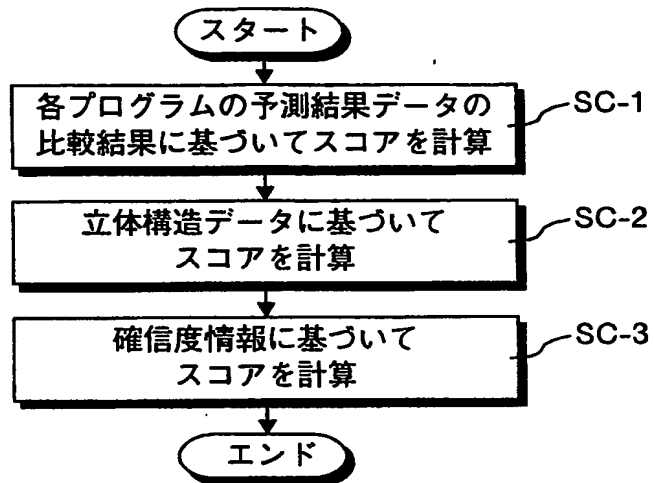


【図 5】

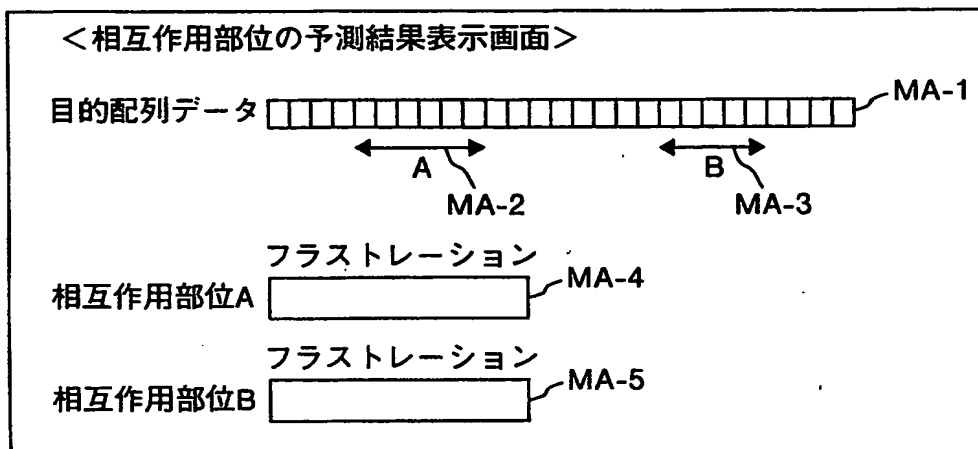


【図6】

(フラストレーション計算処理)



【図 7】



【書類名】 要約書

【要約】

【課題】 タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することを課題とする。

【解決手段】 利用者は、目的のタンパク質の一次配列情報である目的配列データ 1 0 を本発明の相互作用部位予測装置に入力する。タンパク質の一次配列情報から当該タンパク質の断片構造を予測する断片構造予測プログラム 2 0 a ~ 2 0 d に対して入力された目的配列データ 1 0 の断片構造予測シミュレーションを実行する。各断片構造予測プログラム 2 0 a ~ 2 0 d の断片構造予測結果 3 0 a ~ 3 0 d を比較する (6 0) 。この比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算する (7 0) 。計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する (8 0) 。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [500520628]

1. 変更年月日 2000年10月26日

[変更理由] 新規登録

住 所 千葉県千葉市美浜区中瀬1丁目3番地 幕張テクノガーデンD
17

氏 名 セレスター・レキシコ・サイエンシズ株式会社